

Non-parametric estimation of morphological lopsidedness

Nadine Giese^{1,2*}, Thijs van der Hulst¹, Paolo Serra³, Tom Oosterloo^{1,2}

¹*University of Groningen, Kapteyn Astronomical Institute, Landleven 12, NL-9747 AD, Groningen, the Netherlands*

²*ASTRON, the Netherlands Institute for Radio Astronomy, Postbus 2, NL-7990 AA Dwingeloo, the Netherlands*

³*CSIRO Astronomy and Space Science, Australia Telescope National Facility, PO Box 76, Epping, NSW 1710, Australia*

Accepted 2016 June 13. Received 2016 June 12; in original form 2015 December 05

ABSTRACT

Asymmetries in the neutral hydrogen gas distribution and kinematics of galaxies are thought to be indicators for both gas accretion and gas removal processes. These are of fundamental importance for galaxy formation and evolution. Upcoming large blind H I surveys will provide tens of thousands of galaxies for a study of these asymmetries in a proper statistical way. Due to the large number of expected sources and the limited resolution of the majority of objects, detailed modelling is not feasible for most detections. We need fast, automatic and sensitive methods to classify these objects in an objective way. Existing non-parametric methods suffer from effects like the dependence on signal to noise, resolution and inclination. Here we show how to correctly take these effects into account and show ways to estimate the precision of the methods. We will use existing and modelled data to give an outlook on the performance expected for galaxies observed in the various sky surveys planned for e.g. WSRT/APERTIF and ASKAP.

Key words: methods: miscellaneous – galaxies: structure

1 INTRODUCTION

Hydrogen is the most abundant element in the Universe and neutral hydrogen (H I) in particular is of utmost importance in the study of galaxy evolution. H I can extend out to large radii beyond the optical disk, where it can efficiently trace episodes of gas accretion and removal. These include the inflow of gas from the intergalactic medium (IGM), the feedback from the galactic disk through e.g. star formation (“galactic fountain”), the ram pressure exerted by the gaseous medium within which galaxies move, as well as the in-fall of and interaction with gas-rich companions. Observationally, these events can be revealed by the lopsidedness of the H I morphology or kinematics, or by the warping of the H I discs outer regions (Sancisi et al. 2008).

To understand how the above processes affect galaxy evolution we need to assemble large samples of H I detected galaxies and explore the relation between H I asymmetries and other galaxy properties. However, all large, blind H I surveys carried out so far have inadequately large angular resolution for this purpose because they have been done using single dishes (e.g., HIPASS, ALFALFA and EBHIS; see Barnes et al. 2001; Giovanelli et al. 2005; Kerp et al. 2011). In the near future, radio telescopes with higher resolution, adequate sensitivity and high survey speed will for the first time enable this kind of investigations (e.g., the SKA

pathfinders ASKAP and WSRT/APERTIF; see Johnston et al. 2008; Verheijen et al. 2008; Oosterloo, Verheijen, & van Cappellen 2010). The results of these future observing campaigns do not suffer as much from selection effects as previous, smaller surveys (e.g. WHISP, Atlas3D and THINGS; see van der Hulst, van Albada, & Sancisi 2001; Serra et al. 2012b; Walter et al. 2008).

On average 1000 H I detected objects per week are expected for WSRT/APERTIF. The large amount of acquired data can greatly contribute to the study of galaxy formation and evolution but at the same time bring about new challenges with regards to data reduction and analysis. One such challenge is the need to develop analysis methods that are fast, objective, reliable, and sensitive to physical processes that are significant for galaxy evolution.

Existing methods to determine H I properties in galaxies focus on fitting models either to the full H I data cube, the integrated H I map, or the velocity field. For example, TIRIFIC (Józsa et al. 2007) and ^{3D}BAROLO (Di Teodoro & Fraternali 2015) fit 3D tilted-ring models (Warner et al. 1973; Rogstad, Lockhart, & Wright 1974) to data cubes. The possible fit parameters include surface brightness, rotational velocity, scale height, position angle, and inclination at different radii within the galaxy. The rotation curve can also be obtained by fitting the velocity field using the task *rotcur* (Begeman 1987) in the image processing software package GIPSY (van der Hulst et al. 1992). In both the 2D and the 3D cases, asymmetries could either be deter-

* E-mail: giese@astro.rug.nl

mined by comparing the true 3D distribution to a symmetrically fit model or by fitting the galaxies in separate parts, e.g. for the receding and approaching part of the disk and comparing parameters for those. Another existing application for the detection of asymmetries is DiskFit (Spekkens & Sellwood 2007; Sellwood & Sánchez 2010), which can fit non-axisymmetric models to the velocity fields. DiskFit also offers the possibility to model asymmetries in photometric images.

With the large number of galaxies that are expected to be detected within ongoing and future large blind H I surveys, detailed modelling or fitting of every single object will pose problems with respect to execution time, but also resolution of the H I data, since the majority of objects will not have sufficient resolution for detailed fitting. The emphasis here lies on the objects with limited resolution, because only 0.05% of all objects detected in a shallow all-sky survey with WSRT/APERTIF will have sufficient resolution to be fitted with TiRiFiC (see Duffy et al. 2012). Methods using non-parametric approaches include determining asymmetries by harmonically decomposing the integrated H I map (van Eymeren et al. 2011) or determining indices adopted from characterisation in the optical (Holwerda et al. 2011a,b).

In this paper, we re-examine the possibility of using the morphometric parameters employed in Holwerda et al. (2011b) for a non-parametric characterisation of H I in galaxies. In section 2 we will investigate the dependence of different morphometric parameters on data qualities such as noise and angular resolution as well as the inclination of objects. In section 3 we show a possible way to correct the Asymmetry parameter for noise effects and in section 4 we summarise our findings and discuss future work.

2 NON-PARAMETRIC MEASUREMENT OF LOSIDEDNESS

2.1 Measuring morphological shape revisited

Holwerda et al. (2011b) found that using a suitable combination of parameters adopted from optical studies, one can distinguish interacting from non-interacting galaxies. They use a combination of indices measuring Concentration, Asymmetry, Smoothness, Gini, M_{20} and a newly defined parameter Gini- M_{20} (GM) on a subsample of the WHISP survey that had been visually classified in Noordermeer et al. (2005) and Swaters et al. (2002). We will define these indices below. Comparing the visual classifications with their automatically determined parameters, Holwerda et al. (2011b) propose to use either a combination of the Asymmetry parameter and M_{20} or of the Concentration and M_{20} to determine the boundary between interacting and non-interacting galaxies in parameter space. In this Section we revisit their results.

One aspect that deserves attention is the effect of a varying S/N. In particular, the H I data used by Holwerda et al. (2011b) have relatively low S/N, and one may wonder whether this biases the measured parameter values. Consider, for example, the Asymmetry parameter (hereafter also referred to as A or the A parameter), which was first used in

Abraham et al. (1996) and defined in Conselice, Bershadsky, & Jangren (2000) (CBJ00) as:

$$A = \frac{\sum_{i,j} |I(i,j) - I_{180}(i,j)|}{2 \sum_{i,j} |I(i,j)|}, \quad (1)$$

where in the case of H I data I denotes the zero-th moment image of a galaxy cube. I_{180} refers to the same image rotated by 180 degrees. CBJ00 already show that for optical data the measured value of A depends strongly on the S/N of the data even for S/N values of several hundreds. Furthermore, while A can be corrected within 5% of the noise-free value for S/N > 100, no reliable correction is possible below this limit. Lisker (2008) investigated the effect of S/N and aperture size on the determination of the Gini parameter and also found poor performance at low S/N levels. Since most H I data will not reach a S/N level of 100, this aspect needs further attention. We discuss the influence of S/N on the results for H I data in Sect. 2.3.

These noise effects are also relevant for the study of Holwerda et al. (2011b) who use the highest resolution H I WHISP maps for their analysis. The masks (3D segmentations of the cubes that contain signal from the galaxy) have been generated using all voxels above a 1.8σ threshold in the Hanning smoothed 60" data cube. The 2D H I images were created as the zero-th moment of the masked version of the highest resolution data cube. To check the applicability of the A parameter for data cubes of quality comparable to that of the WHISP ones, we estimated the mean and maximum S/N per galaxy of the full galaxy sample in an aperture containing 90% of the galaxy flux. To this end we estimated the S/N along a line of sight using only the voxels in the mask:

$$(S/N)_{i,j} = \frac{\sum_k I_{i,j,k}}{\sqrt{N_{i,j}} \sigma}. \quad (2)$$

Here, $I_{i,j,k}$ are the voxels that are included in the object mask, with the index k referring to the velocity/frequency axis. $N_{i,j}$ is the number of channels that have been added to the pixel with coordinates (i,j) in the integrated H I image. σ is the RMS noise in the cube. We determine the global S/N for a galaxy as the average of the line of sight S/N distribution:

$$S/N = \langle (S/N)_{i,j} \rangle. \quad (3)$$

We find that the average as well as the maximum signal to noise is always significantly below the level of 100. Fig. 1 shows the distribution of the mean and maximum S/N values throughout the sample. Therein, the majority of galaxies have a mean S/N below 10 and the maximum pixel values are below 100 for all galaxies. Thus, none of the objects have sufficient signal to noise to use the Asymmetry parameter without suffering from noise effects. In Holwerda et al. (2011b) no attempt was made to correct the A values for noise.

We used the same WHISP maps of the 141 galaxies analysed in Holwerda et al. (2011b) to measure the parameters mentioned above. We chose the optical centres provided by Noordermeer et al. (2005) and Swaters et al. (2002) for the calculation of Concentration, Asymmetry, M_{20} and GM . We excluded the 10 galaxies, for which Holwerda et

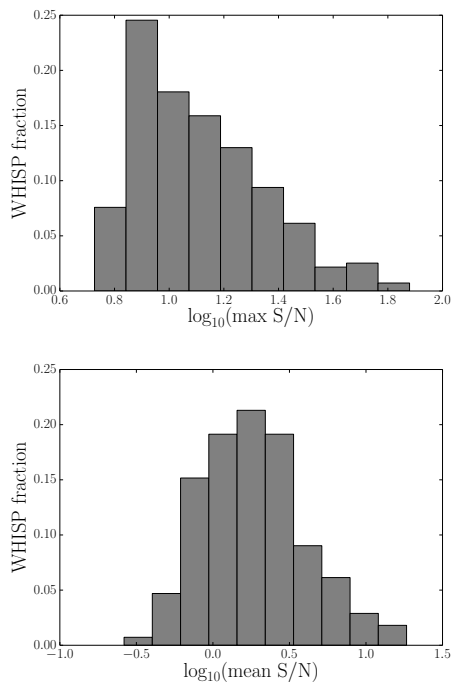


Figure 1. Histograms of the mean and maximum pixel value in the WHISP high resolution signal to noise maps - For the majority of objects the mean signal to noise in the moment maps is below 10. The maximum signal to noise is below 100 in all cases.

al. (2011b) did not calculate any morphometric parameters. Additionally, we excluded UGC 10448 since the H I observation has been carried out for the wrong target (see Noordermeer 2006). This left us with 130 objects. We used a 1σ threshold to select the pixels that are used for calculation. This threshold is low enough to include the outer disk of the sample galaxies. Holwerda et al. (2011b), too, apply a threshold to the WHISP H I images in order to measure the morphometric parameters. Their adopted threshold is, however, in H I column density (the actual value is not given but we note that Holwerda et al. (2011a) adopt a threshold of $3 \times 10^{19} \text{ cm}^{-2}$ when analysing the THINGS sample). We prefer to use a S/N threshold because noise is the main driver of systematics in the measurement of these parameters. The actual choice of S/N threshold is not trivial, because it can have a significant effect on the calculated parameters. If a galaxy breaks up into several pieces because of a too high threshold or because the S/N is too low, the Asymmetry parameter will reach very high values close to the maximum, although a visual inspection might reveal that the object is symmetric. On the other hand, if the threshold is chosen too low, the Asymmetry value will reach higher values as well, due to the strong effects that noise has on pixels with low flux.

Our values for the A parameter already show a strong deviation from the results in Holwerda et al. (2011b) (see Fig. 2). There is a large number of galaxies for which we measure a lower A . The horizontal error bars are taken from the tables A1 and A2 in the online version from Holwerda et al. (2011b). There is no further description of these errors with the online material. Thus, considering the error

estimation in Holwerda et al. (2014), we assume that the errors given in the online tables are the uncertainties stemming from the error in the central position as well as the flux values in the total H I images. For the error bars in Fig. 2 we took the square root of the sum of the squared errors. The vertical errors are generated in a similar way. The component that results from the uncertainty in the central position has been estimated following Holwerda et al. (2014) who vary the centre within a Gaussian with an FWHM of three pixels.

For a simple illustration of why we obtain lower values we selected three galaxies which have the highest possible A parameter according to the analysis in Holwerda et al. (2011b) and for which we estimate a significantly lower A . These galaxies are UGC 798, UGC 4458 and UGC 4666. Their A values have been marked with special symbols in Fig. 2. In Fig. 3 we show the moment zero map (left panel) as well as the absolute difference between the moment zero map and its 180 degrees rotated version (right panel) for all three galaxies. The intensities in both panels are on the same scale for each of the three galaxies. Taking equation 1 into account, the panels showing the difference must have an integrated intensity which is twice as high as for the moment zero map in order for the galaxy to have the maximum A of 1. However, this is not the case for any of the examples.

We suspect that the errors in A are due to the use of the wrong optical centres for the objects, since this would explain not only the discrepancy in the Asymmetry but also other parameters like Concentration, M_{20} and GM, which will be defined below. In particular, the extreme values of $A = 1$ could be explained with the centre of the galaxy being located outside of the mask defining the galaxy. We tried to reproduce the Asymmetry results from Holwerda et al. (2011b) by using the H I flux weighted centre instead of the optical, swapping the pixel coordinates of RA and Dec, assuming the wrong Epoch while converting from RA and Dec to pixel coordinates, not excluding other objects that are in the same integrated H I image and selecting a centre at random. However, none of the approaches reproduces the Asymmetry values consistently. An explanation for the extreme A values can be given for a few galaxies which have H I images using mJy as a unit for the density rather than Westerbork Units ($1 \text{ W.U.} = 5 \text{ mJy}$). In these cases the thresholds are too high because the wrong unit is assumed leaving zero pixels for the evaluation of the Asymmetry parameter and hence giving A values of 1. UGC 798 and UGC 4666, which are shown in Fig. 3, are among those galaxies.

In addition to the Asymmetry parameter, Holwerda et al. (2011b) included the parameters Concentration, Smoothness, Gini, M_{20} , and the newly defined parameter GM. The Concentration parameter is usually estimated as the logarithm of the ratio between the radii containing 80% and 20% of the total flux of an object (Bershady, Jangren, & Conselice 2000):

$$C = 5 \log \frac{r_{80}}{r_{20}} \quad (4)$$

Since the H I gas distribution in galaxies does in general have a lower central concentration than e.g. the stellar counterpart, the Concentration parameter estimated from total H I images will have lower values than the ones estimated from optical images. Holwerda, Pirzkal, & Heiner (2012) point out

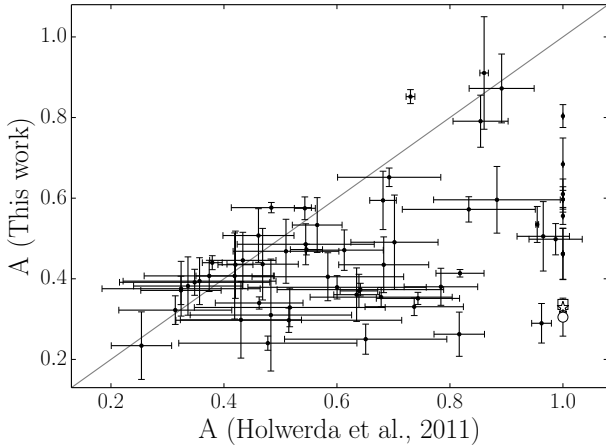


Figure 2. Comparison of the values for the Asymmetry parameter for 141 WHISP galaxies. For the comparison we use a table with parameters provided with the online version of [Holwerda et al. \(2011b\)](#). Since the A values in that table cover the range $[0, 2]$, we divided them by 2. The vertical error bars have been estimated as the square root over the sum of the squared uncertainties resulting from the error in the central position of the galaxy and the errors in the flux values of the total H I image. The horizontal error bars are calculated similarly with the values for the errors taken from the tables A1 and A2 of the online version of [Holwerda et al. \(2011b\)](#). The values estimated in [Holwerda et al. \(2011b\)](#) are higher or approximately equal to the values that we estimate. We show the moment zero images as well as the differences for the rotated images in Fig. 3 for the examples that have been marked with the square, star and open circle symbols. We could not determine a systematic effect which could have led to the significantly higher values in [Holwerda et al. \(2011b\)](#).

that there had been an error in the Concentration estimation in a previous paper. This should be taken into account when evaluating the results presented in [Holwerda et al. \(2011b\)](#).

The Smoothness parameter is defined as the difference between the original image and the smoothed version and has been developed by [Conselice \(2003\)](#) as a clumpiness parameter to evaluate the degree of small scale structures. In this work we only estimate the Smoothness parameter for the WHISP galaxies in order to compare with [Holwerda et al. \(2011b\)](#). However, since the choice of smoothing kernel necessary for its measurement is arbitrary, we exclude this parameter from further analysis.

The Gini parameter, which was first used on astronomical data by [Abraham, van den Bergh, & Nair \(2003\)](#), has its origin in economics, where it is used to describe the inequality of the distribution of wealth among a nation's citizens. Expressed in the context of images it is defined as the mean of the absolute difference between all combinations of pixels intensities I_i . Its limits are zero, if all pixels have the same intensity, and one, if all intensity is contained in one pixel. The Gini parameter is defined as

$$\text{Gini} = \frac{1}{\bar{I}n(n-1)} \sum_i (2i - n - 1)I_i, \quad (5)$$

where \bar{I} is the mean pixel intensity and n the number of pixels. The values I_i have to be sorted in an ascending order.

In addition to the Gini parameter, [Lotz, Primack, & Madau \(2004\)](#) introduce the moment of light parameter M_{20} .

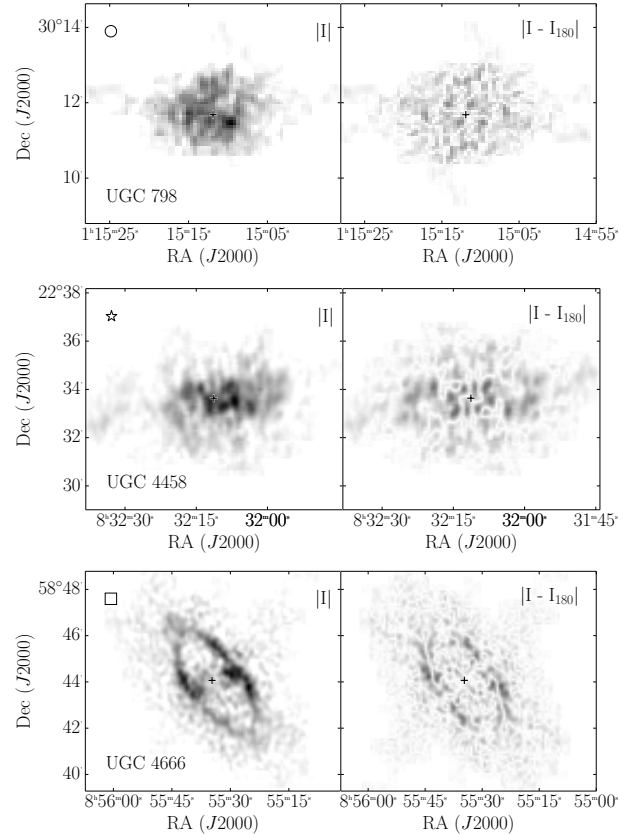


Figure 3. Moment zero images (left panels) and difference images (right panel) for the moment zero image and its 180 degree rotated version. The black crosses indicate the optical centres of the galaxies. The Asymmetry values measured in [Holwerda et al. \(2011b\)](#) for all three examples have a maximum of 1, which indicates that the galaxies are highly asymmetric. The difference does not show signs for strong asymmetries.

For every pixel, a second order moment is defined as the product of the pixel intensity and the squared distance from the centre of the galaxy:

$$M_i = I_i [(x_i - x_c)^2 + (y_i - y_c)^2]. \quad (6)$$

The M_{20} parameter is calculated as the fractional sum over the second order moments of the brightest pixels containing 20% of the total intensity and the sum over all second order moments:

$$M_{20} = \log_{10} \frac{\sum_i M_i}{M_{tot}}, \text{ where } \sum_i I_i < 0.2 I_{tot}. \quad (7)$$

It traces bright structures, especially if they are offset from the centre. In optical images the M_{20} parameter is therefore sensitive to structures like spiral arms, bars and phenomena associated with merger processes, like e.g. multiple nuclei. Additionally, [Holwerda et al. \(2011b\)](#) introduce another parameter, which is a combination of the Gini and the M_{20} parameters, called GM :

$$GM = \frac{1}{\bar{M}n(n-1)} \sum_i (2i - n - 1)M_i. \quad (8)$$

Comparing this definition to equations (5) and (6), GM is the Gini parameter for the distribution of second order moments M_i .

Table 1. Number and fraction of interacting WHISP galaxies from the combined Noordermeer et al. (2005) and Swaters et al. (2002) set that have been correctly identified as interacting using the respective criterion.

Selection criterion	number (fraction) of galaxies
(a) $GM > 0.56$	5 (18%)
(b) $A > 0.71$	6 (21%)
(c) $A > -0.15 \times M_{20} + 0.54$	7 (25%)
(d) $M_{20} < 13 \times GM - 8.7$	7 (25%)
(e) $A > -0.6 \times GM + 1.01$	7 (25%)
(f) $C < 13 \times A - 6.5$	8 (29%)
(g) (a)∨(b)∨(c)∨(d)	10 (36%)

We estimated all parameters for the WHISP data set to see if the results from Holwerda et al. (2011b) can be confirmed. We used the visual classifications from Noordermeer et al. (2005) and Swaters et al. (2002) to divide the combined sample into interacting and non-interacting objects. It should be noted that galaxies that are classified as non-interacting do not show any sign of interactions. This, however, does not proof that they are not interacting. For convenience and to be consistent with the Holwerda et al. (2011b) notation, we will use the term non-interacting for galaxies that do not show any signs of interaction. Fig. 4 shows the distribution of morphometric parameter values for these two different groups. We find that a number of interacting galaxies can be roughly separated from the non-interacting sample using $A > 0.65$ or $GM > 0.6$. However, this separation is of a different nature compared with the findings in Holwerda et al. (2011b). We find that the distribution of the Asymmetry parameter values for the interacting sample spans a range that exceeds the range of the non-interacting sample. The GM parameter shows a similar, but less pronounced feature. The results from Holwerda et al. (2011b) do not show these distinct parameter ranges. However, they find a much clearer separation between the peaks of the two distributions in the Asymmetry as well as GM parameter.

Fig. 5 shows 2D projections of the distribution of the interacting and non-interacting subsamples in the space defined by all morphometric parameters (see Scarlata et al. 2007 for the first example of such plot). In this space we can define a number of criteria to find interacting galaxies without including any non-interacting objects. The criteria from Table 1 are displayed as dashed lines in Fig. 5.

Although an individual classification returns better results for the Noordermeer et al. (2005) sample, we chose to use the combined Noordermeer et al. (2005) and Swaters et al. (2002) sample to cover a large range of galaxy properties and data qualities. A combination of the Asymmetry, GM and M_{20} parameters is sufficient to recover $\sim 36\%$ of the galaxies that have been visually classified as interacting without including any of the non-interacting objects. A principle component analysis helps making the separation between the objects that can be classified as interacting in the original parameter space from the remaining galaxies more pronounced. The increase in the number of galaxies that can be correctly classified as interacting is, however, not significant.

Another interesting aspect for the use of these parameters is the classification of lopsidedness in galaxies (please

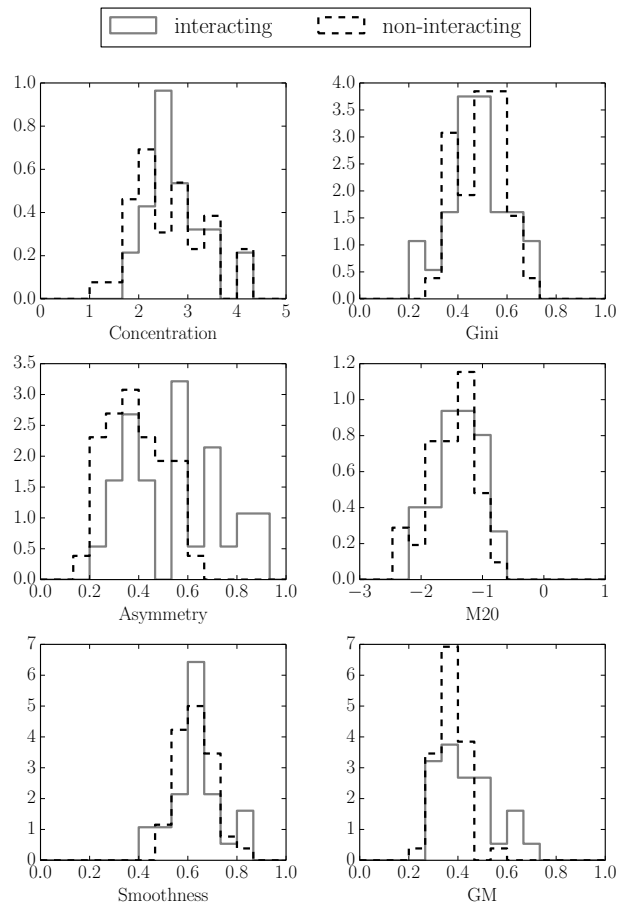


Figure 4. Concentration, Asymmetry, Smoothness, Gini, M_{20} and GM for the visually inspected WHISP subsamples from Noordermeer et al. (2005) and Swaters et al. (2002) - There is a small trend in the Asymmetry parameter with the distribution of interacting galaxies reaching higher values than non-interacting galaxies.

see Jog & Combes (2009) for an overview of the lopsidedness phenomenon). Since Noordermeer et al. (2005) and Swaters et al. (2002) use different scales for the strength of lopsidedness, we chose to analyse both sets separately. The results are shown in Fig. 6, 7 and 8.

There is no clear separation between the non-lopsided and the lopsided groups in the histograms. The multi parameter plots with the more detailed groups of lopsidedness (Fig. 7 and 8) feature some strongly lopsided galaxies that are separated from the other groups in Asymmetry- GM space and in their higher GM and Asymmetry values. The other groups of lopsidedness, however, show a very strong overlap among each other. Overall these trends are not sufficient to classify galaxies into different categories of lopsidedness. This might be attributed to the problem that different data properties like the signal to noise and the resolution strongly influence the values of the morphometric parameters, such as Asymmetry.

An important problem in estimating the Asymmetry parameter becomes apparent when determining it for a galaxy at both different resolutions and different signal to noise ratios. The two effects are closely connected through the increase in signal to noise per pixel when smoothing and

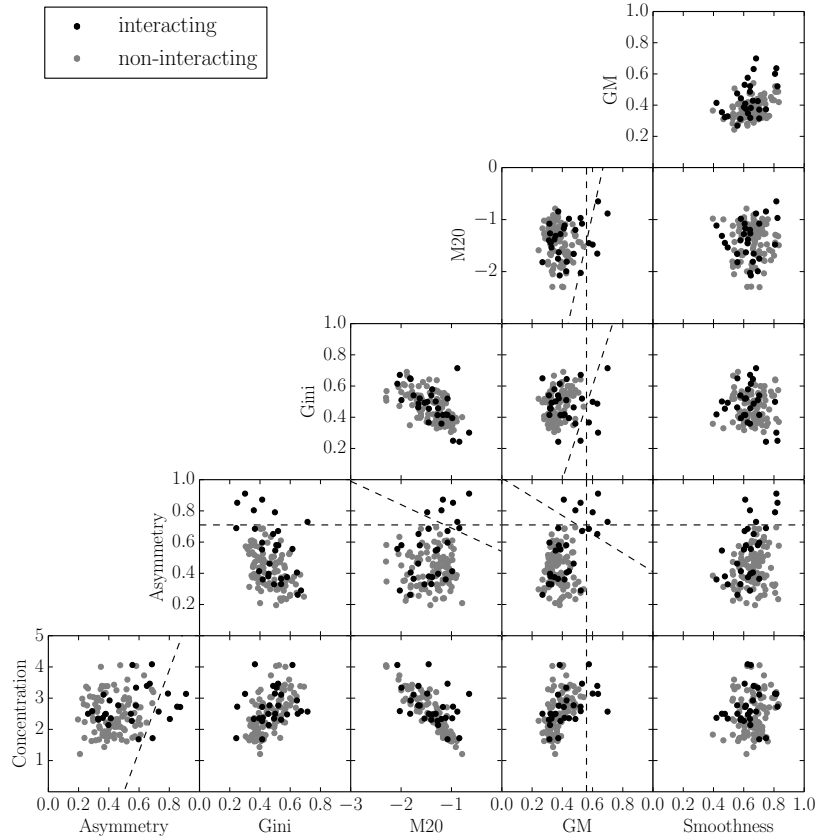


Figure 5. Concentration, Asymmetry, Gini, M_{20} , GM and Smoothness for the visually inspected WHISP subsamples from [Noordermeer et al. \(2005\)](#) and [Swaters et al. \(2002\)](#) - We do not find the same trends in parameter space as indicated in [Holwerda et al. \(2011b\)](#).

thereby lowering the resolution of an integrated H I image. Using the three different resolution versions ($12''/\sin(\delta)$, $30''$ and $60''$) of galaxies from the WHISP survey, we noticed that with lower resolution the measured asymmetry decreases. An illustration of this effect is shown in Fig. 9. We picked UGC 2953 as an example of a low signal to noise case at highest resolution. We estimated the asymmetry using the optical centre as the reference point. The result for the asymmetry parameter changes from highly asymmetric ($A = 0.68$) in the full resolution image to mildly/intermediately asymmetric ($A = 0.22$) in the lowest resolution image. The latter value agrees better with the visual classification as non-lopsided by [Noordermeer et al. \(2005\)](#). This example also illustrates that the dynamic range in current observations of H I disks is typically a few orders of magnitude less than in optical disks, affecting the sensitivity of the A parameter.

In the following sections we aim to show how certain image and galaxy properties influence the outcome of the measurement of the parameters mentioned above. To investigate the effects of signal to noise, resolution and inclination on the parameters used in [Holwerda et al. \(2011b\)](#) in more detail, we generated models using the tilted ring fitting code TiRiFiC ([Józsa et al. 2007](#)), in which we could vary the mentioned properties independently.

2.2 Model galaxies

We used the tilted ring fitting code TiRiFiC to generate models of galaxies with varying intrinsic morphological asymmetry, inclination, and size relative to the angular resolution of the data. Input parameters include circular velocity, surface brightness profile, position angle, inclination and scale height for every simulated ring. We adopted the surface brightness profiles from [Martinsson \(2011\)](#)

$$\Sigma_{\text{H I}}(R) = \Sigma_{\text{H I}}^{\text{max}} \cdot e^{-\frac{(R - R_{\Sigma, \text{max}})^2}{2\sigma_{\Sigma}^2}}, \quad (9)$$

where $R_{\Sigma, \text{max}} = 0.39R_{\text{H I}}$ denotes the position of the peak, $\sigma_{\Sigma} = 0.35R_{\text{H I}}$ the width, $\Sigma_{\text{H I}}^{\text{max}}$ the peak of the profile, and $R_{\text{H I}}$ the H I radius. We will use $R_{\text{H I}}$ as the H I scale length in units of modelled rings for the surface brightness profile and refer to it as the H I radius hereafter. We chose the rotation curve representative for intermediate mass galaxies (see Fig. A1).

To introduce morphological lopsidedness (asymmetries in the moment images), we added a harmonic surface-brightness distortion of first order to the models. These can be realized in TiRiFiC by choosing an amplitude and a phase for the distortion for every ring. An example of TiRiFiC input parameters can be found in Fig. A1. The last two panels represent the values for the amplitude (SM1A) and the phase (SM1P) of the first order harmonic surface-brightness distortion. The values for the amplitude are cho-

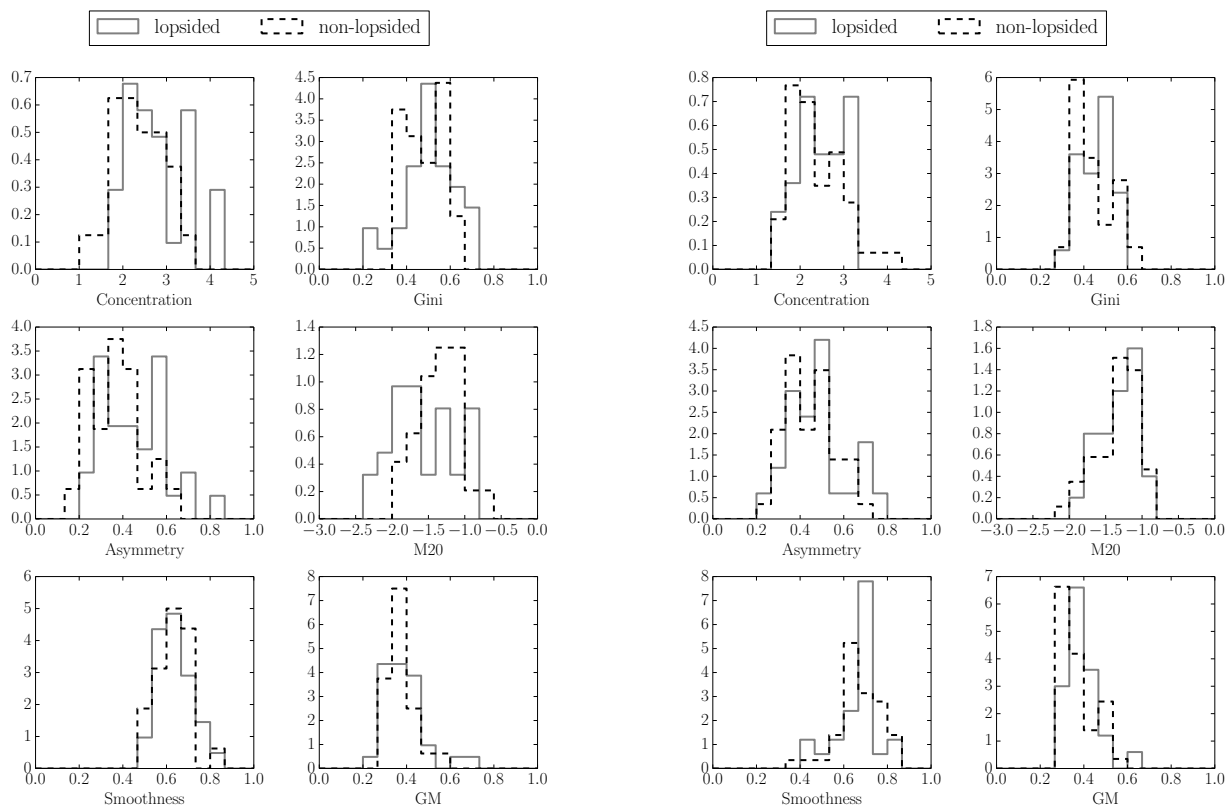


Figure 6. Concentration, Asymmetry, Smoothness, Gini, M_{20} and GM for the visually inspected WHISP subsamples from Noordermeer et al. (2005) (left panel) and Swaters et al. (2002) (right panel).

sen to be close to the surface brightness profile within the inner rings (except the innermost one). The strength of the added lopsidedness is then defined as a multiplier to the SM1A profile with a value of 0 representing symmetric objects and a value of 1 for strongly lopsided models. In the latter case all flux in the rings that are affected is azimuthally shifted using the phase of the harmonic surface-brightness distortion. An example for the different lopsidedness levels is shown in Fig. 10. Using this definition, the strongest lopsided galaxy does not have a maximum Asymmetry value of 1. This value is, however, usually not reached for real galaxies, where the Asymmetry values usually span a range between 0 and 0.6.

All our initial models consist of 11 rings. For the variation of the resolution of our models, we generated high resolution models, where the beam size corresponded to the sampling size. To achieve a certain angular galaxy size and hence resolution in beams across the major axis we first re-gridded these models to the desired angular size and then “observed” them by applying the instrumental broadening function. Using this approach we assumed that our highest resolution models corresponded to the clean components of a real observation. The resulting cubes were passed to the source finding application SoFIA (Serra et al. 2015) to determine the mask of every galaxy, i.e. all pixels that contain signal. The method in SoFIA used to define pixels that contain signal is the smooth and clip method. The 3D data cube is smoothed in all 3 dimensions at different scales and all pixels that lie above a 4σ noise threshold at that par-

ticular resolution are added to the mask of the cube. The mask is a separate 3D cube of the same size as the cube containing the galaxy, with all pixels belonging to the galaxy having the value 1 and all other pixels being 0. The smooth and clip method is an established way of defining the extent of the H I emission in galaxies and searches for emission at different scales, because it assumes that the signal to noise is highest when the smoothing kernel matches the scale of the emission. For the experiments done in the next few subsections masks were generated from the noiseless cubes and later used on all versions with varying noise level. Here, we assumed, that with low signal to noise faint emission that is not detected within the actual source finding step can be recovered by growing the masks in a subsequent step. This is of course an idealised approach, which fails if the signal to noise in the 3D cube becomes too low. We chose to take this approach nonetheless to be able to investigate the effect of increasing noise level without sections of the galaxy being cut off. Additionally, we did not choose a signal to noise threshold to restrict the pixels that entered into the calculation. We added noise using a noise cube that was generated from the Stokes Q component of an observation with the WSRT (see Serra et al. 2012a). The noise cube has a resolution of $30''$. Consequently, all our TIRIFIC models are smoothed to a round beam of $30''$.

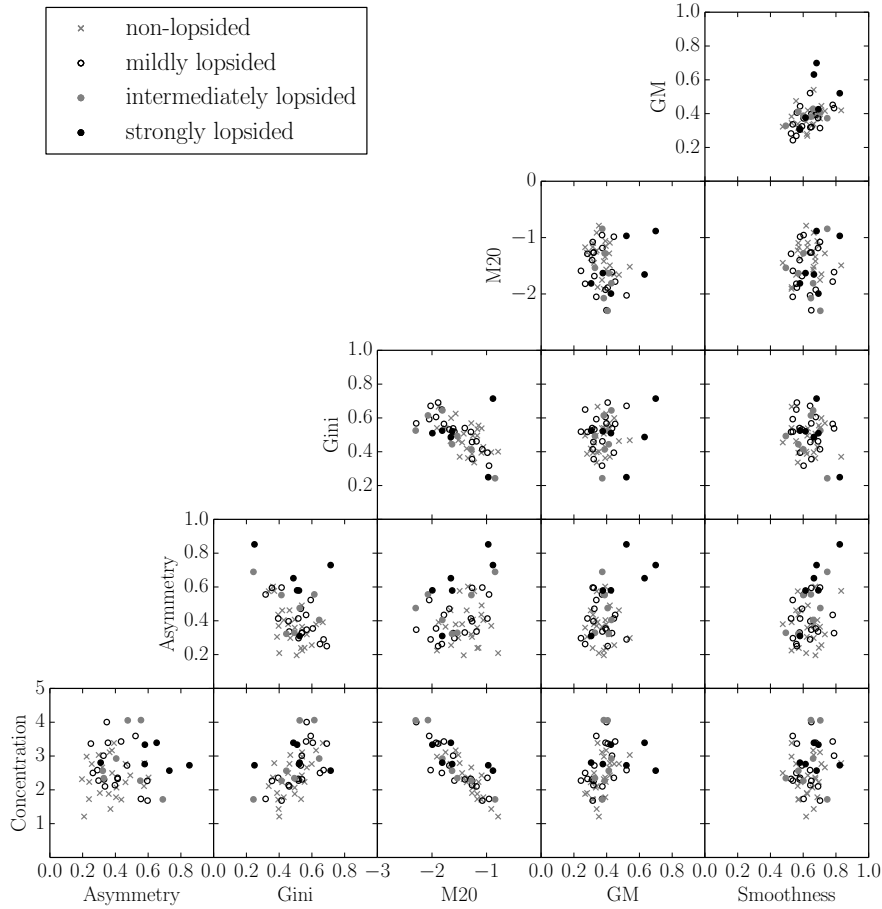


Figure 7. Concentration, Asymmetry, Smoothness, Gini, M_{20} and GM for the visually inspected WHISP subsamples from Noordermeer et al. (2005).

2.3 Effect of S/N variations

To investigate the influence of signal to noise on the recovered parameters C , A , G , M_{20} , and GM we generated noiseless model cubes for five galaxies with the same radial surface brightness density profile (see Fig. A1) and different intrinsic lopsidedness due to increasing amplitudes for the harmonic surface brightness distortion (see Fig. 10). We chose a resolution of 5 beams across the major axis. This choice is based on the results from Duffy et al. (2012). Using simulations, they found that only 0.5% of the galaxies expected in a shallow survey with Westerbork/APERTIF would have a resolution of > 5 beams across the major axis. Fig. 11 shows an example of one such galaxy which was generated using the parameters shown in Fig. A1 and with different levels of noise added. For decreasing signal to noise the perceived lopsidedness is more and more determined by the noise instead of the intrinsic lopsidedness.

Apart from the varying noise levels and the amplitude of the harmonic surface brightness distortion we kept all other galaxy properties, in particular the resolution and inclination, constant (see caption in Fig. 12). The results for the signal to noise dependence of the morphometric parameters can be found in the first column of Fig. 12. The dotted line indicates the maximum signal to noise in the WHISP data set. Galaxies observed with Westerbork/APERTIF or

ASKAP will have similar S/N characteristics. The different shades of grey represent the degree of lopsidedness - with light grey for non-lopsided and black for highly lopsided galaxies.

It becomes immediately apparent that the Asymmetry parameter (second row) is the most promising parameter with regards to the separation of objects with different degrees of lopsidedness. For S/N values above 100 the recovered A does not deviate significantly from the intrinsic value. All examples are well separated with more lopsided galaxies having higher A values. Towards lower signal to noise values the measured A increases. At a S/N of about 3 the two galaxies with the lowest lopsidedness show similar A values while the three intermediate to high lopsidedness cases can still be separated very well. For even lower S/N the separation between all models becomes very small. The discovered trends show that it is not possible to compare A values for galaxies with different S/N below 100. It is however noteworthy, that the Asymmetry parameter always deviates towards higher values. Furthermore, the slow and smooth increase for lower S/N values suggests that a correction of this noise bias is possible. We will present an approach for the correction in section 3.

For S/N values above 100 the other parameters show very similar trends. Except for small variations they remain at the initial values for infinite signal to noise. The Con-

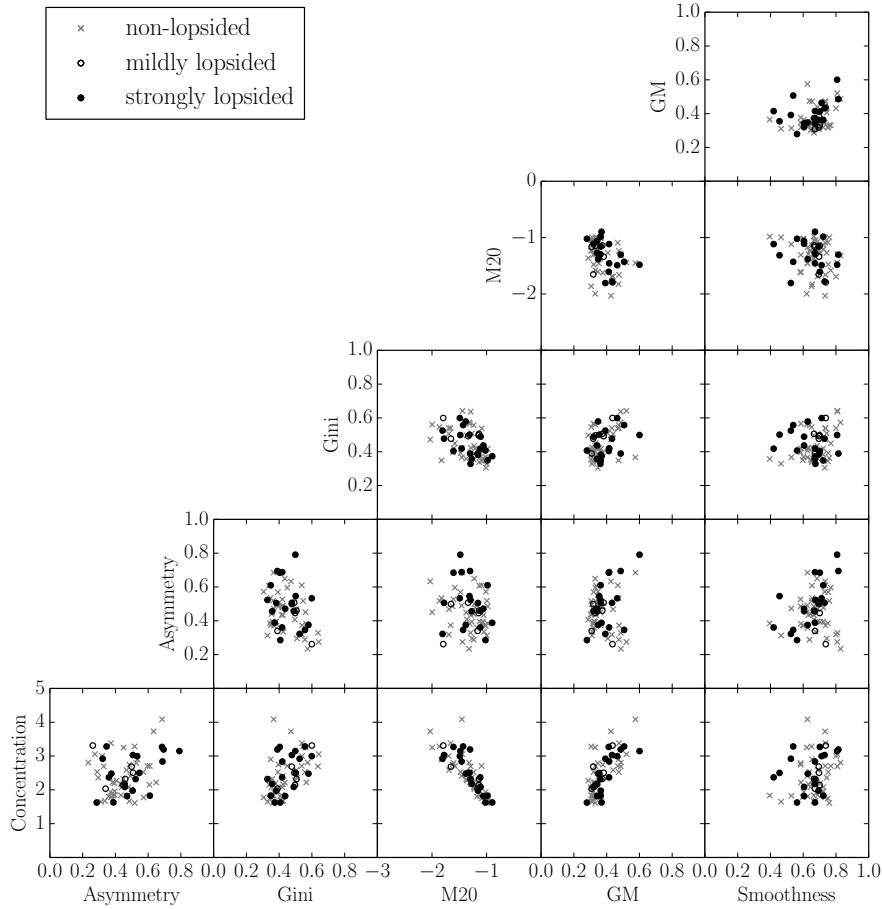


Figure 8. Concentration, Asymmetry, Smoothness, Gini, M_{20} and GM for the visually inspected WHISP subsamples from [Swaters et al. \(2002\)](#).

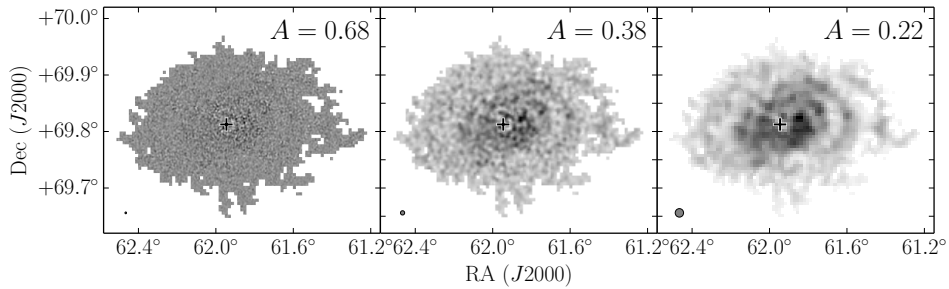


Figure 9. The WHISP galaxy UGC 2953 at 3 resolutions (10.0''x11.2'', 27.0'', 54.2''). The black crosses indicate the optical galaxy center. The highest resolution image of this galaxy has a high asymmetry parameter although it is visually classified as intermediately asymmetric. With lower resolution the signal to noise increases and the asymmetry parameter decreases.

centration parameter does not show any significant separation between galaxies with different degrees of lopsidedness. This can be attributed to the way lopsidedness was introduced. The harmonic surface brightness distortion works on individual rings, thus the total intensity in annuli is conserved and the Concentration parameter remains unchanged regardless of the amplitude of the distortion. The Concentration parameter therefore does not seem suited to separate galaxies with different degrees of lopsidedness.

The Gini parameter shows at best a marginal separation between the four strongest lopsided models for S/N values

above 100. For values below 100 the Gini value decreases as expected because with higher noise level the noise is dominating the flux values in each pixel and the distribution of the flux values is not restricted to the pixels containing the strongest signal of the galaxy.

The M_{20} parameter does not show any meaningful separation of the 5 different lopsidedness models. It is not possible to distinguish between the intermediate and high asymmetry models. Only the lowest asymmetry example is well separated at lower M_{20} values. For S/N values below 1 this trend however also disappears.

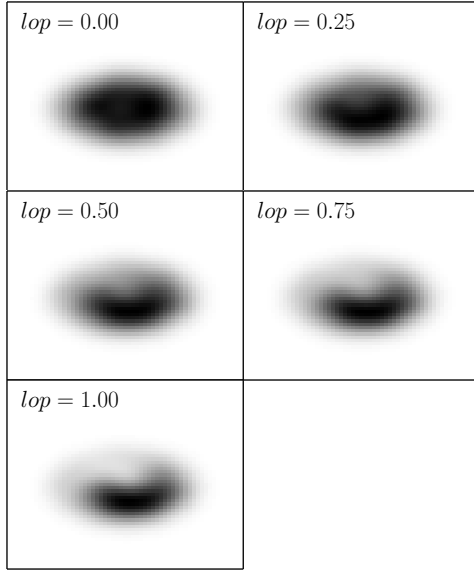


Figure 10. Example of a TiRiFiC model galaxy with different levels of lopsidedness. An example set of input values for these galaxies are shown in Fig. A1.

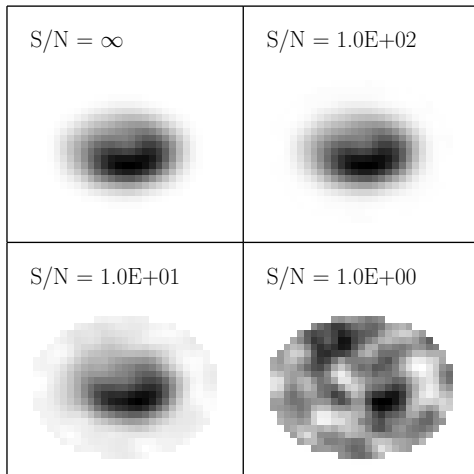


Figure 11. Example of a TiRiFiC model galaxy with decreasing signal to noise level (S/N according to Eq. 2 and 3). The input values for this galaxies are shown in Fig. A1. With decreasing S/N level the apparent lopsidedness is more and more replaced by asymmetric structures in the noise.

For high S/N values the GM parameter exhibits a trend, where higher values correspond to higher lopsidedness in the model. However, already at S/N values above 100, the measured GM parameter slightly decreases with decreasing S/N . For S/N values of the order of the WHISP galaxies models start to overlap in covered GM ranges.

Based on Fig. 12 A is the most useful parameter for the separation of galaxies with different levels of lopsidedness. It is superior due to the smooth trends with signal to noise and should be preferred provided that it is possible to correct the values for the positive offset, which we will refer to as the noise bias.

2.4 Effect of resolution

We used the TiRiFiC models described above to investigate the influence of angular resolution on the estimation of the morphometric parameters. In this step we did not add noise to the models in order to allow us to separate the two effects. Fig. 13 shows an example that has been set to resolutions from 20 to 3 beams across the major axis. We degrade the resolution of our models in terms of beam sizes across the major axis by re-gridding them on different pixel scales before convolving with a Gaussian beam.

Decreasing the resolution of the galaxies is the same as placing the models at larger distances. In Fig. 12 (second column) we show the results for the same galaxy examples as used in subsection 2.3. As before, the darker shades of grey correspond to higher intrinsic lopsidedness.

The asymmetry parameter again turns out to separate the different lopsidedness models best. For lower resolution and thus examples at larger distance the measured A value decreases compared to the intrinsic one. Down to a resolution of about three beams across the major axis all models are still separable, although models with lower degrees of lopsidedness start covering similar value ranges. Since a decrease of resolution is connected to a loss of information, it is not possible to correct the estimated A for this effect. It is thus advisable to group galaxies into different classes of angular resolution for a comparison of the degree of lopsidedness among different objects. Another possibility is the decrease of the angular resolution of objects in a group to the value of the object with the lowest resolution. This would at the same time increase the S/N , which in turn decreases the influence of the noise bias on the A parameter. For the large blind surveys that are planned for WSRT/APERTIF or ASKAP this approach is however not the best, considering that the majority of objects will have very low resolution (≤ 3 beams across the major axis).

The Concentration parameter does neither show any significant separation between the different objects nor any trend with resolution in the measured value. Gini and GM do show some separation between the models with the highest degree of lopsidedness for the highest resolutions (20 and 10 beams across the major axis). Towards lower resolutions this separation however decreases. The models with intermediate to low degrees of lopsidedness do not show any significant separation. The M_{20} parameter does not consistently separate objects with different degrees of lopsidedness over the range of resolutions investigated.

2.5 Effect of inclination

For the investigation of the effect of inclination on the morphometric parameters we used the same model approach as described in Sect. 2.3 and 2.4. For this experiment we fixed the resolution at five beams across the major axis and chose an infinite signal to noise. All other galaxy properties are the same as in the previous sections. The only parameter that was varied was the inclination. Fig. 14 shows four galaxies at different inclinations varying from 0° to 90° . With increasing inclination the apparent lopsidedness of the models appears to decrease.

For a detailed analysis we again used the same models with five different degrees of lopsidedness. The results are

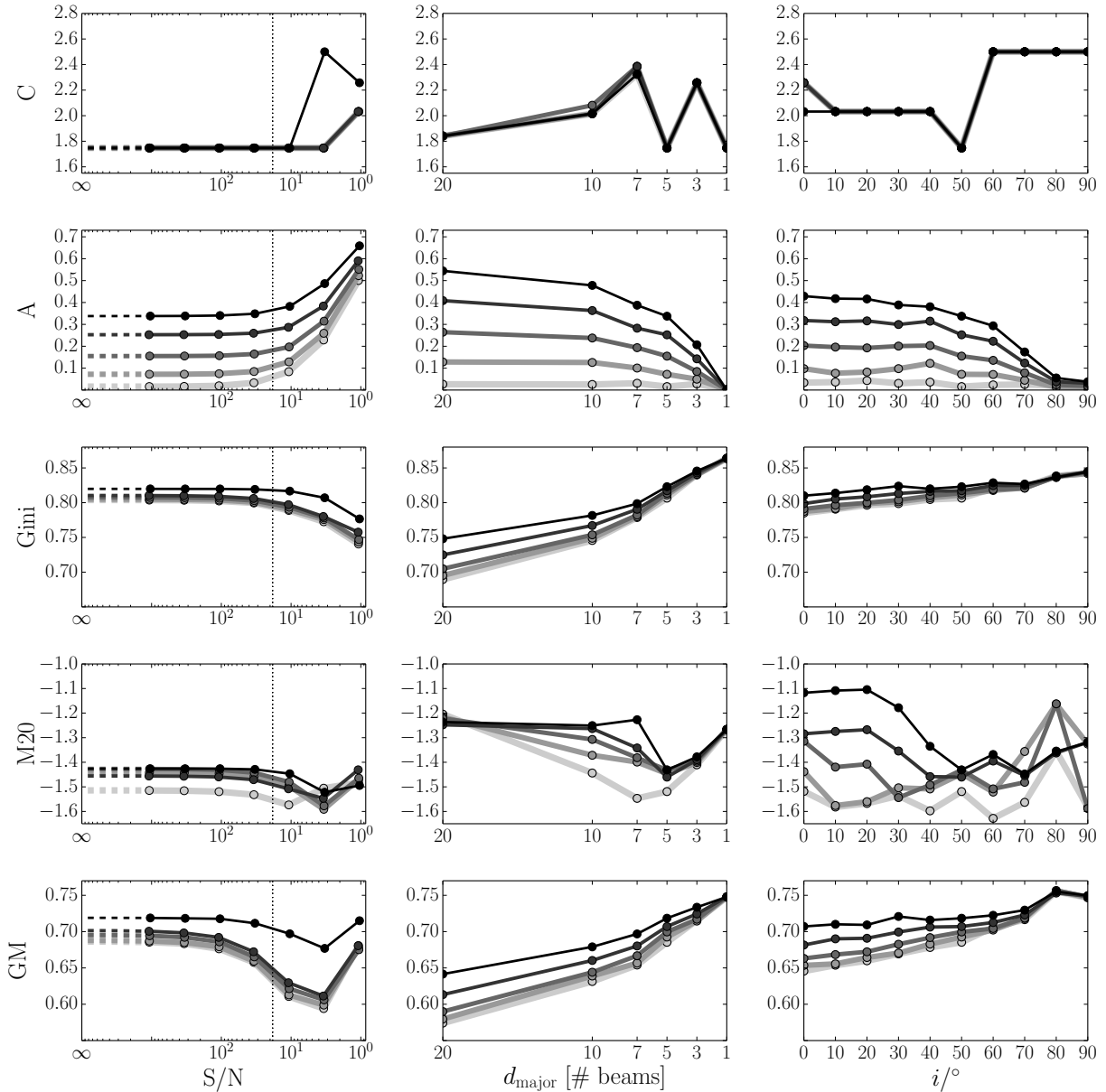


Figure 12. Signal-to-noise, resolution and inclination dependence of C , A , Gini, M_{20} and GM for galaxy models with different intrinsic morphological lopsidedness. The grey level is indicating the strength of the lopsidedness with light grey for non-lopsided to black for highly lopsided models. The five different levels correspond to the 5 examples shown in Fig. 10. All models were based on 11 rings. For every panel only one galaxy parameter was varied respectively (S/N , d_{major} and i). The non-varying parameters chosen for the models were $S/N = \infty$, $i = 50^\circ$, $d_{\text{major}} = 5$ beams. The dotted line indicates the maximum signal to noise for galaxies in the WHISP sample (Eq. 3). The most suitable parameter for the estimation of the degree of lopsidedness in a galaxy is the A parameter. It shows a very good separation between the different lopsidedness models and features the slowest rate of change.

shown in the third column of Fig. 12. When comparing the last two columns in Fig. 12 it seems that inclination and resolution have similar effects on the Asymmetry, the Gini and the GM parameters except for a smoother trend for the resolution. This is not surprising as the change towards higher inclination is mainly a decrease of resolution along the minor axis of a galaxy. The parameter separating the different lopsidedness models best is again the Asymmetry parameter. All other parameters have similar limitations to the ones found in Sect. 2.4.

Our findings for the Asymmetry parameter differ from

results from previous work. Holwerda et al. (2011a) find that below an inclination of 80° the measured A does not suffer from inclination effects. In contrast, using TIRIFIC models we find that (1) the inclination limit for reliable A measurement is lower and (2) the effect of inclination on the measured A strongly depends on the azimuthal position of the lopsidedness within the galaxy.

Holwerda et al. (2011a) use existing integrated H I images from the THINGS survey to investigate the dependence of A on inclination. They rotated the images around the major axis to simulate galaxies with higher inclination and

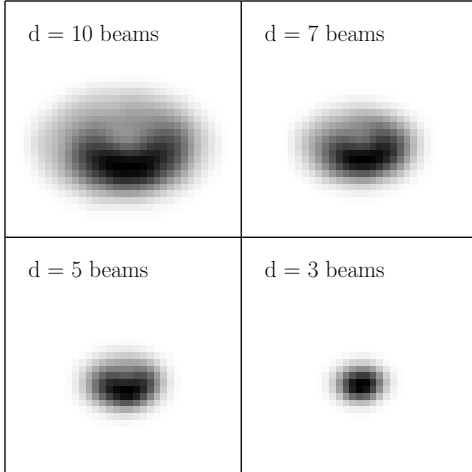


Figure 13. Example of a TIRIFIC model galaxy with decreasing angular resolution. The input values for this galaxies are shown in Fig. A1. With decreasing angular resolution the apparent lopsidedness decreases with respect to the intrinsic one.

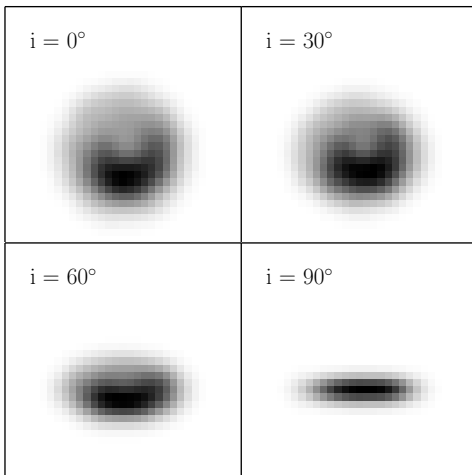


Figure 14. Example of a TIRIFIC model galaxy with decreasing angular resolution. The input values for this galaxies are shown in Fig. A1. With decreasing angular resolution the apparent lopsidedness decreases with respect to the intrinsic one.

rebinned the images to the initial pixel size. Thereby they assumed that the galaxies featured an infinitely thin disk. This assumption however does not hold for the majority of galaxies. To demonstrate the difference between this 2D and a full 3D approach, we generated TIRIFIC model galaxies with increasing inclination and degree of lopsidedness. The azimuthal position of the lopsidedness was chosen at an intermediate position between major and minor axis. For the sake of comparability with Holwerda et al. (2011a), we chose an angular resolution of 20 beams across the major axis for our models.

For every galaxy we calculated the Asymmetry parameter from the zero-th moment map. Additionally, we used the moment maps at zero inclination and rotated them around the major axis as described in Holwerda et al. (2011a). The results are shown in Fig. 15. The 2D approach does not show

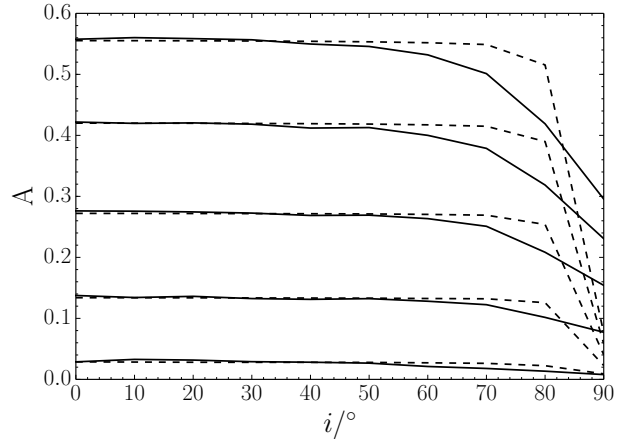


Figure 15. Comparison of recovered A changing with inclination. The solid/dashed line pairs represent models with different levels of intrinsic asymmetry (corresponding to the five models represented in Fig. 10). The dashed lines represent a simple rotation of the galaxy moment zero image around the major axis as presented in Holwerda et al. (2011a). The solid lines show the trends for A in full 3D models which have been set to increasing inclinations. The full 3D model approach already shows a decrease in recovered A at inclinations of about 50-60 degrees.

a significant change in recovered A up to an inclination of 80 degrees. The full 3D approach already shows a decrease of the recovered A for lower inclinations. The difference can be explained by the thickness of the disk. The assumption of an infinitely thin disk in the 2D approach does not hold for real galaxies. Inclination limits should thus be derived from full 3D models.

Another important effect that is directly connected to the inclination is the position of the lopsidedness within the galaxy. In our models we changed the azimuthal position of the lopsidedness within the galaxy and calculated the A parameter from the zero-th moment map generated from the full 3D cube. In Fig. 16 we show the results for two models with low (grey) and high (black) degree of lopsidedness. The different curves for each galaxy represent models with lopsidedness at different azimuthal positions between the major and minor axis. The curves that feature a steeper slope at high inclinations represent the models where the lopsidedness is positioned along the minor axis of the galaxy. For models with lopsidedness along the major axis the decrease of the recovered A towards high inclinations is more shallow.

In general, one can conclude that the influence of inclination on the A estimation already becomes significant at around 50 – 60°. Therefore, lopsidedness recovered from galaxies at higher inclinations should be seen as a lower limit or excluded from statistical analyses when using this definition for the Asymmetry parameter.

2.6 Asymmetry is the preferred parameter

Summarising our findings from the last sections, we find that the Asymmetry parameter is the most useful morphometric parameter for the separation of galaxies with different degrees of lopsidedness. We also find a signal to noise limit of ~ 100 below which the measured Asymmetry parameter deviates significantly from the intrinsic one. It is still possi-

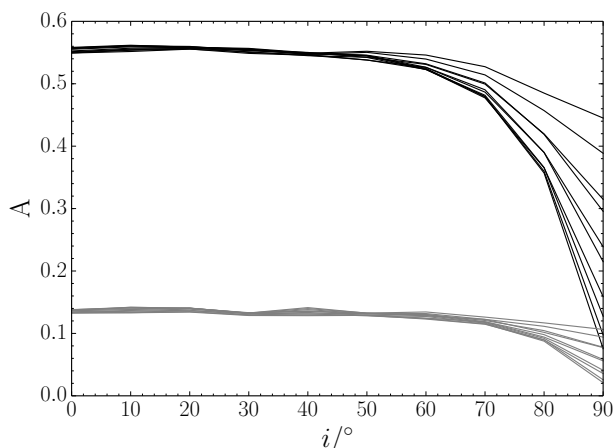


Figure 16. Noiseless TIRFiC models with increasing inclination and different azimuthal position of lopsidedness in the galaxy disk. The light grey and black lines represent two models with low and high intrinsic morphological lopsidedness (corresponding to the $lop = 0.25$ and $lop = 1.0$ models in Fig. 10). The different lines of the same grey scale level stem from the same galaxy model. The only difference is the azimuthal position of the source of lopsidedness within the galaxy disk. The measured A decreases with increasing inclination. The steepness of the decrease depends on the azimuthal position of the lopsidedness in the galaxy. Models with azimuthal positions aligned with the major axis feature a less steep decrease than models with lopsidedness aligned with the minor axis.

ble to separate objects with different degrees of lopsidedness down to a resolution of 3 beams across the major axis. Additionally, we find an inclination limit of about 60° above which the decrease in the measured Asymmetry parameter depends strongly on the azimuthal position of the source of the lopsidedness within the disk.

In a next step we used our models to investigate further how well the parameters can (alone and in combination with each other) distinguish between objects of different morphological lopsidedness. We restricted the model galaxies to three versions of lopsidedness: non-lopsided, intermediate and highly lopsided (corresponding to lopsidedness multipliers of 0, 0.5 and 1, see Fig. 10). Furthermore, we adjusted the properties to the findings from the previous sections and chose inclinations from 0 to 60 degrees, resolutions in beams across the major axis from 3 to 20 and signal to noise levels above 100. We estimated the parameters Asymmetry, Concentration, Gini, M_{20} and GM as defined in Holwerda et al. (2011b).

The results are shown in Fig. 17. Therein, all panels showing the combination of parameters which include the Asymmetry parameter show a pronounced separation between lopsided and non-lopsided galaxies. In contrast, the strongest overlaps occur in combinations with the Concentration parameter and in the Gini- GM combination. The M_{20} parameter also shows overlaps with other parameters. In the combinations with Gini and GM these overlaps still allow for a separation between certain fractions of the non-lopsided and strongly lopsided objects.

Fig. 18 shows the separation of the different lopsidedness groups when only single parameters are used. Obviously, in the presence of high signal to noise, the A param-

eter is the most promising one with regards to the separation between galaxies with different degrees of lopsidedness. For all other parameters, every group with a particular degree of lopsidedness spans the complete range of possible values.

Considering the good separation of the three different model groups it is reasonable to prefer the Asymmetry parameter to identify galaxies with lopsided distributions, if the resolution and inclination of the objects are within the determined limits and the A parameter can be corrected for the noise bias.

3 CORRECTION OF ASYMMETRY VALUES AT LOW S/N

The A parameter turns out to be the most suitable parameter for the evaluation of the degree of lopsidedness within the total H I image of a galaxy. It shows the smoothest and least steep trends in the dependence on the different data qualities or galaxy properties. Furthermore, in the presence of high signal to noise it can be used independent from the other parameters to separate lopsided from non-lopsided objects. In this section we investigate whether its measured value can be corrected to take into account the effect of various data characteristics.

The effects of inclination and resolution cannot be corrected for because both are due to a loss of information in the data. However, these effects can be taken into account by considering the measured A as a lower limit of the intrinsic A (the underlying Asymmetry value without noise). Galaxies within groups of the same resolution and inclination could still be evaluated with respect to each other. In contrast, the effect of noise is additive and in combination with the aforementioned effects makes an interpretation of recovered A values complicated if not impossible.

We define the positive offset, the noise bias, which becomes significant for galaxies with mean signal to noise lower than 100, as follows:

$$A_{\text{measured}} = A_{\text{intrinsic}} + \text{Bias}. \quad (10)$$

A very simple approach to avoid the noise bias is to increase the signal to noise in the moment maps. This can be achieved by e.g. smoothing or de-noising in Fourier space. However, this can either lead to a loss of information or to an introduction of new patterns which are not real. CBJ00 presented an approach for the prediction of the noise bias. Therein, the mask defining the area of the galaxy is shifted around in areas outside the mask to estimate the numerator of the asymmetry only in regions which contain noise (and background). The corrected A parameter is defined as:

$$A_{\text{corrected, CBJ00}} = \frac{\sum |I - I_{180}|}{2 \sum |I|} - \frac{\sum |B - B_{180}|}{2 \sum |I|}, \quad (11)$$

where B has the same shape as I (using the same mask), but defines a region outside of the galaxy that contains background as well as noise. The minuend on the right hand side of Eq. 11 corresponds to the Bias in Eq. 10. Taking into account that H I data ideally has a zero background level, this approach would in general lead to an overcorrection for

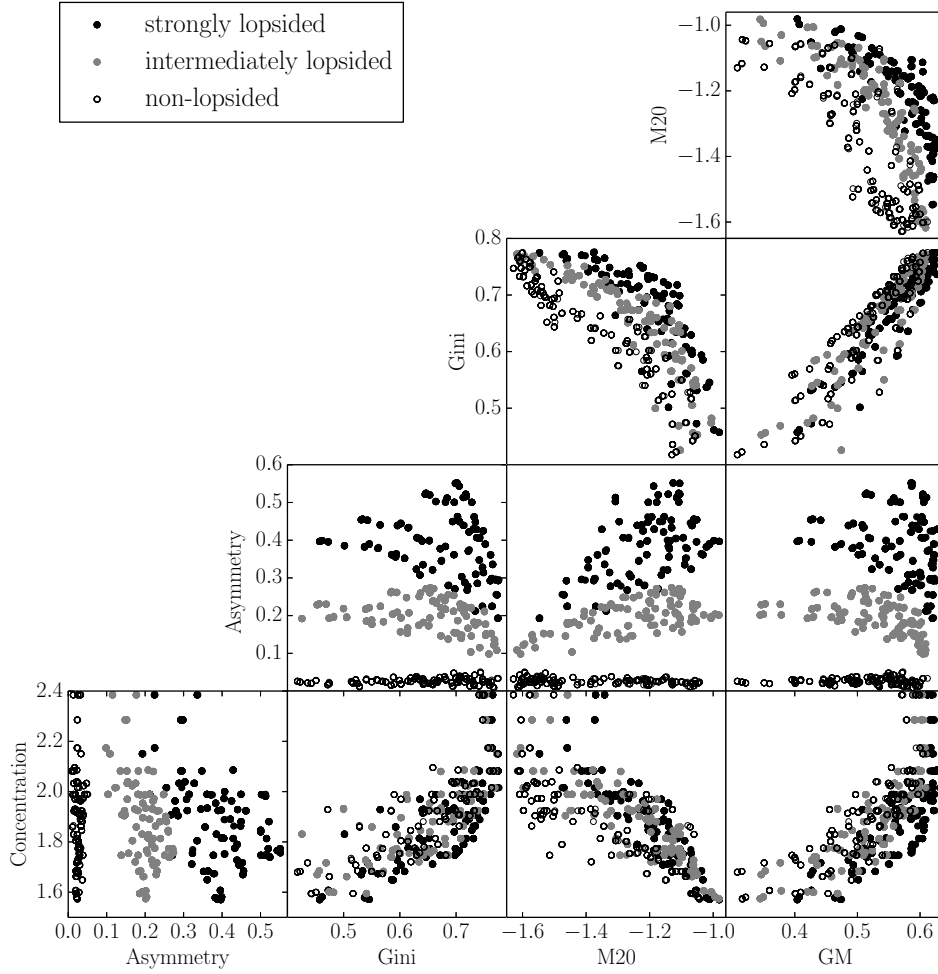


Figure 17. Correlations between morphometric parameters measured on TIRIFIC model galaxies with varying signal to noise, H I radius, resolution, and inclination, taking into account the limitations determined in sections 2.3, 2.4 and 2.5. We restricted the models to three of the lopsidedness categories illustrated in Fig. 10: strongly lopsided ($lop=1.0$), intermediately lopsided ($lop=0.5$) and non-lopsided ($lop=0$). The Asymmetry parameter in combination with all other parameters clearly separates lopsided from non-lopsided galaxies, whereas there are slight to strong overlaps between the remaining parameters.

a non-zero noise level. To illustrate this point, consider the following:

$$\begin{aligned}
 & \sum |I - I_{180}| - \sum |B - B_{180}| \\
 &= \sum |(I_0 + n) - (I_{0,180} + n_{180})| - \sum |n - n_{180}| \\
 &\leq \sum |I_0 - I_{0,180}|
 \end{aligned} \tag{12}$$

where $I_{0,(180)}$ are the pixel intensities of the image without noise and $n_{(,180)}$ samples of Gaussian noise (and their respective 180 degree rotated versions); the sum of which constitute the components of the moment map ($I = I_0 + n$). It becomes apparent that the bias corrected Asymmetry parameter always represents a lower limit for the intrinsic A :

$$\Rightarrow A_{\text{intrinsic}} \geq A_{\text{measured}} - \text{Bias}(\text{CBJ00}) = A_{\text{corrected,CBJ00}} \tag{13}$$

Moreover, it should be noted that this approach does not take into account the fact that the bias has a larger influence on pixels with low intensity. It is therefore not suitable as a correction for the noise bias in H I data.

A new approach that we propose here is to use model galaxies to approximate the noise bias. To investigate the feasibility of this approach, we built a data base of TIRIFIC models (set S1) changing different properties such as the noise level (hence the signal to noise), the size of the galaxy relative to the angular resolution, the inclination, the intensity and azimuthal position of the morphological lopsidedness, and the H I radius. We again use the surface brightness definition from Martinsson (2011) and the rotation curve displayed in Fig. A1. The lopsidedness is introduced as a first order harmonic surface brightness distortion. We used SOFIA to determine the pixels that contain emission. This time we used the cubes including noise, with the result that some galaxies could not be found due to low S/N. For our models we estimated not only the noise bias, but also values that could be determined from observational data, e.g. the mean signal to noise, the total flux, the A value (which includes the bias) and the number of 2D and 3D pixels in the mask, which defines the region of the galaxy. As a simple test, we created two additional sets of galaxies, where we varied the input galaxy properties (e.g. inclination, resolu-

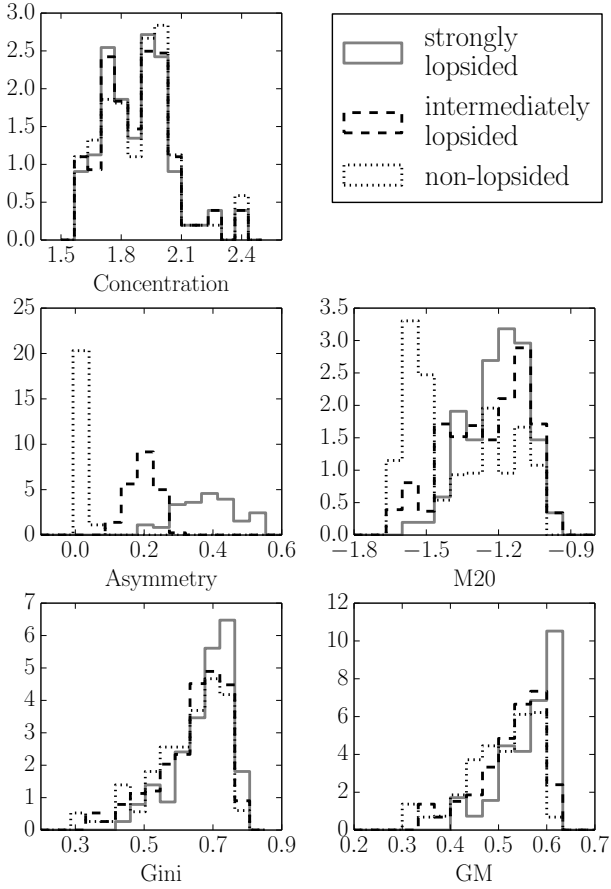


Figure 18. Morphological parameters measured on TiRiFiC model galaxies with varying signal to noise, H I radius, resolution, and inclination, taking into account the limitations determined in sections 2.3, 2.4 and 2.5. We restricted the models to three of the lopsidedness categories illustrated in Fig. 10: strongly lopsided ($lop=1.0$), intermediately lopsided ($lop=0.5$) and non-lopsided ($lop=0$). The Asymmetry parameter clearly separates lopsided from non-lopsided galaxies.

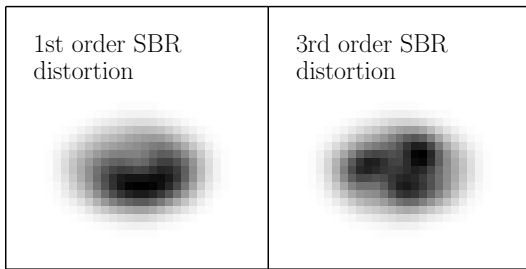


Figure 19. Example galaxies with lopsidedness introduced by a first order (left panel) and a third order (right panel) harmonic surface brightness distortion.

tion, etc.) in the same range covered by the objects of set S1. The lopsidedness in the first set (S2) was generated by a first order harmonic surface-brightness distortion while in the second set (S3) lopsidedness was generated by a third order surface-brightness distortion (see Fig. 19 for an example).

Table 2. Model parameters for the different galaxy model sets

Parameters	Galaxy set S1	Galaxy set S2	Galaxy set S3
Inclinations [°]	0,10,20,...,90	15, 35, 55	
Azimuthal angle of asymmetry [°]	0,10,20,...,90	15, 45	
Lopsidedness multiplier	0, 0.25, 0.5, 0.75, 1.0	0.15, 0.45, 0.75	
Order of harmonic distortion	1st	1st	3rd
$R_{H I}$ in rings	1, 3, 5, 7, 9, 11	4, 8, 10	
$d_{\text{galaxy beams}}$ [#]	20, 15, 10, 7, 5, 3, 1	12, 8, 6, 4	
Noise amplitude multiplier	0.5E-8, 5E-7, 5E-6, 5E-5, 1.58E-5, 5E-4, 5E-3, 5E-2, 5E-1, 1.58E-1, 1.58, 5, 50	3E-4, 3E-3, 3E-2, 3E-1, 3, 30	

The different parameters of the model sets are listed in Table 2.

Since we know that the bias is significant in particular for low pixel differences between the H I image and its rotated version we define another parameter, which measures the signal to noise of the intensity difference S/N_{Δ} :

$$S/N_{\Delta} = \left\langle \frac{|I(i, j) - I_{180}(i, j)|}{\sqrt{\frac{1}{2}(N(i, j) + N_{180}(i, j))\sigma}} \right\rangle. \quad (14)$$

I and I_{180} are the pixel intensities, N and N_{180} the number of channels along the line of sight through the cube that were added up to the pixel value I and I_{180} respectively, and σ is the RMS noise in the cube. If we plot the bias for the model galaxies of set S1 as a function of the measured A parameter, the mean signal to noise and the newly defined S/N_{Δ} we find a smooth trend (see Fig. 20), which suggests that these three parameters are well suited for estimating the bias in galaxy A measurements.

We estimated the A parameters for all the galaxies in our training (S1) and test sets (S2, S3) with and without noise. The results for the two test sets can be seen in the upper panels of Fig. 21. We plotted the difference between the measured and the intrinsic A parameter against the measured one. The measured A is clearly influenced by the noise bias, hence the positive offset. The colours of the points are indicators of each galaxy's signal to noise, which is applied on a logarithmic scale. We only plot galaxies that have a S/N below 100, since A values do not need correction for S/N values above 100. As expected, the models with lower mean signal to noise feature a higher bias value.

In a next step we used the machine learning library SCIKIT-LEARN¹ and the first set of galaxy models to interpolate the noise bias for the galaxies in the test sets. For the bias interpolation we used a 3D parameter space, which

¹ <http://scikit-learn.org>

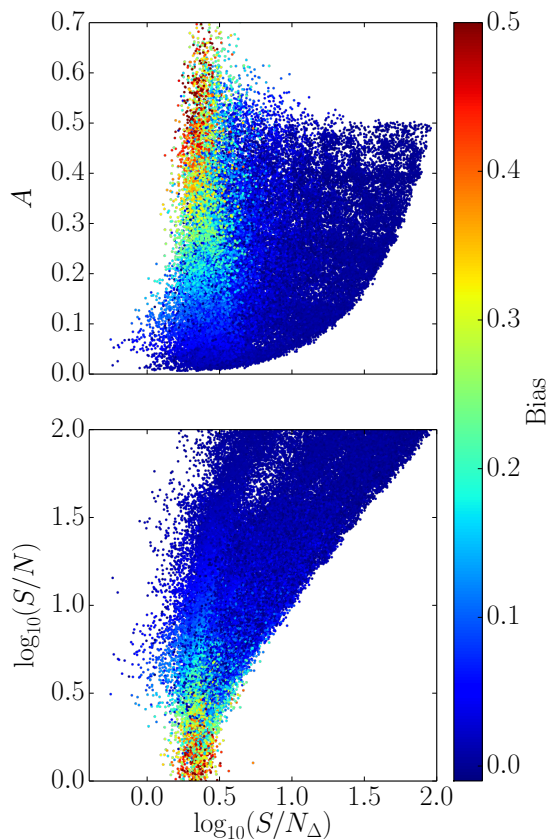


Figure 20. Bias dependence of the galaxies in the model data base S1 on the morphometric A parameter, the mean S/N and the S/N_{Δ} parameter. We find a smooth trend with higher bias values concentrating in regions of high A values, low signal to noise and low S/N_{Δ} . The smooth trend makes these three parameters good candidates for a noise bias correction.

consists of the new parameter S/N_{Δ} as defined in Eq. (14), the mean signal to noise (see Eq. 2 and 3) and the measured A parameter of the galaxies including noise. These parameters can be measured on a pixel base and do not require any further knowledge about the galaxy structure. Since our model set S1 is large enough to slow down any algorithm significantly, we divide the parameter space into subspaces and correct the test sets by interpolating in these smaller parameter regions. This speeds up the process notably because some parameter regions do not have to be explored due to the absence of test examples in these regions. Furthermore, we excluded all test examples that turned out to have a negative total flux assuming that these sources would not have been found using the source finder.

We use the machine learning suit SCIKIT-LEARN with the galaxy set S1 as the training set. The trained algorithm is then used to predict the bias values for the test sets S2 and S3, which are subtracted from the estimated A parameter for the galaxies with noise. The difference between the corrected A value and the intrinsic one is shown in the lower panels of Fig. 21. The correction works quite well for both data sets. Offsets from the intrinsic values spread symmetrically around zero. This is particularly surprising since the model set S1 and test set S3 have been generated using different forms of lopsidedness.

For a comparison we also estimated the bias using the prescription given in CBJ00. We shifted the masks around the galaxies only varying the spatial coordinates. We chose shift positions that span a 3 by 3 grid with a grid spacing equal to the galaxy size in RA and Dec and the actual galaxy position being in the central grid point. This way we end up with eight bias measurements for a galaxy. The final bias is estimated as the median of this bias sample distribution. The results are shown in Fig. 22.

As expected, the resulting Asymmetry parameter values are over-corrected. But as mentioned earlier, this particular correction method is not suited for H I data.

For the model based correction method we further investigated how well the A parameter can be bias corrected within the parameter space that was used for correction. We show the results in Fig. 23 using the same two projections as in Fig. 20. We find that there is no clear trend in the correction quality in the shown parameter space. Most galaxies can be corrected within 0.05 of the original value. There are a few outliers in the galaxy set S2 which have higher deviations with values for the difference between intrinsic and estimated A up to 0.33.

These results lead us to the conclusion that the presented method is suitable for e.g. a simple classification into non-lopsided, intermediately lopsided and strongly lopsided objects, as it is indeed possible to correct the noise bias to a certain extent in a very simple way without the need for additional knowledge about any properties of the galaxy.

4 CONCLUSIONS AND OUTLOOK

We found that the application of the optical parameters Concentration, Asymmetry, Gini, M_{20} and GM adopted and defined in Holwerda et al. (2011b) on H I data does not produce meaningful results unless effects like the signal to noise of the data as well as the inclination and resolution of the galaxies are taken into account. We found that above a signal to noise limit of about 100, the optical parameters can be used on H I data without correction. For a mean signal to noise in the H I maps below this threshold, it is necessary to apply a correction.

We have found the Asymmetry parameter to be the most useful among the set of investigated parameters for measuring galaxy lopsidedness. For low signal to noise the Asymmetry parameter features a noise bias, which is always positive. Inclination and resolution effects influence the result in a very similar way. The intrinsic A can be recovered to a certain degree down to a resolution of about 3 beams across the major axis. Since the resolution as well as the inclination effects cause a loss of detail, it is not possible to recover the full information about the lopsidedness of a galaxy. The recovered A parameter should always be seen as a lower limit. When comparing galaxy samples they should therefore be sorted into categories of similar resolutions and inclinations. We found an inclination limit of approximately 60° above which the steepness of the decrease of A depends strongly on the azimuthal position of the lopsidedness within the disk. For lopsidedness aligned with the minor axis, this decrease is steeper than for the ones aligned with the major axis.

For the correction of the effects of low S/N ($\ll 100$)

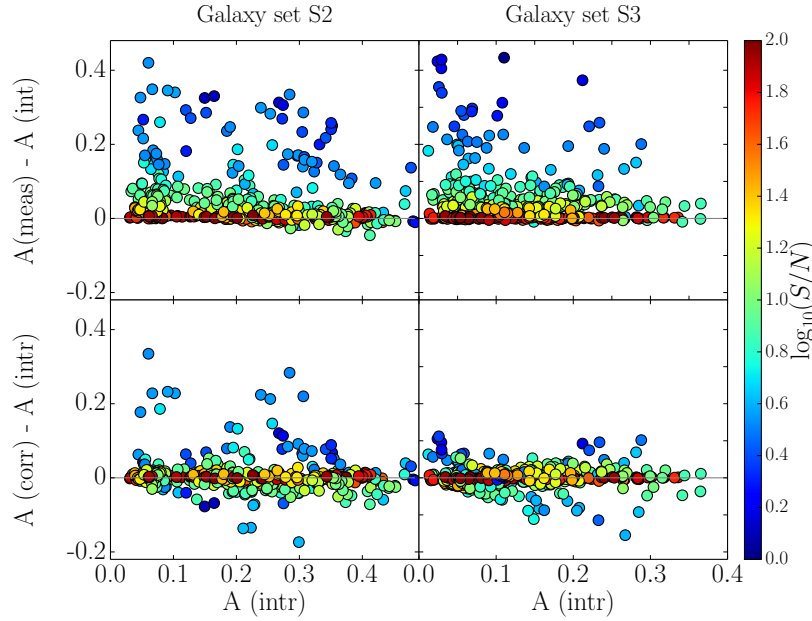


Figure 21. Intrinsic, measured and corrected A values for the galaxy model sets with first order (S2) and third order (S3) harmonic surface brightness distortions. Both corrections result in a better agreement between the intrinsic and the recovered A values.

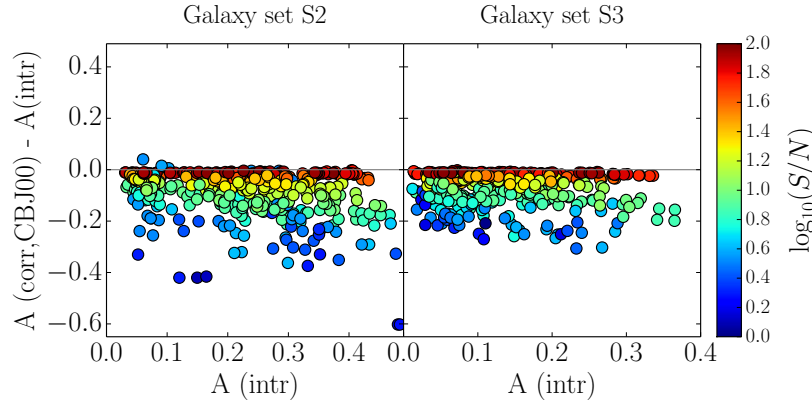


Figure 22. Corrected A values for the galaxy model sets with first order (S2) and third order (S3) harmonic surface brightness distortions using the bias correction from [Conselice, Bershadsky, & Jangren \(2000\)](#). As expected the Asymmetry parameter values are over-corrected.

we have investigated the approach of using a galaxy model data base to find the noise bias as a function of S/N and the recovered Asymmetry parameter. We find that it is possible to correct the noise bias, provided that the galaxy models in the correction set are similar to the galaxies that need bias correction. Our models are all similar in the shape of the surface brightness profile and the rotation curve. The model data base should be updated with different versions of surface brightness profiles and rotation curves. One could consider building a data base using the properties determined from data of real objects that have been classified visually. Another approach that might be considered is the use of real galaxies with high signal to noise. These could be used as templates to investigate the effects of resolution and signal to noise. However, since only a small number of observed galaxies that have been observed are in the signal to noise regime necessary (≈ 100), they can only be seen as an addition to models if they are used for bias correction. The

machine learning tool SCIKIT-LEARN is a very useful tool to interpolate the bias for objects using a pre-built data base. Since it is mostly only necessary to classify different categories of asymmetry or lopsidedness (e.g. strong, intermediate, no lopsidedness), one could consider using a discrete classification scheme, which would eventually also offer more options to employ more sophisticated machine learning techniques, such as neural networks to classify galaxies.

In a next step the methods should be tested on real data. An ideal test case is the WHISP survey galaxy sample. A significant number of galaxies has been classified visually ([Swaters et al. 2002](#); [Noordermeer et al. 2005](#)) and can be used as a test set for the verification of the method. This is the subject of a subsequent study. A scientific example for the application of the relation between asymmetry or morphology and environment density of galaxies using surveys such as

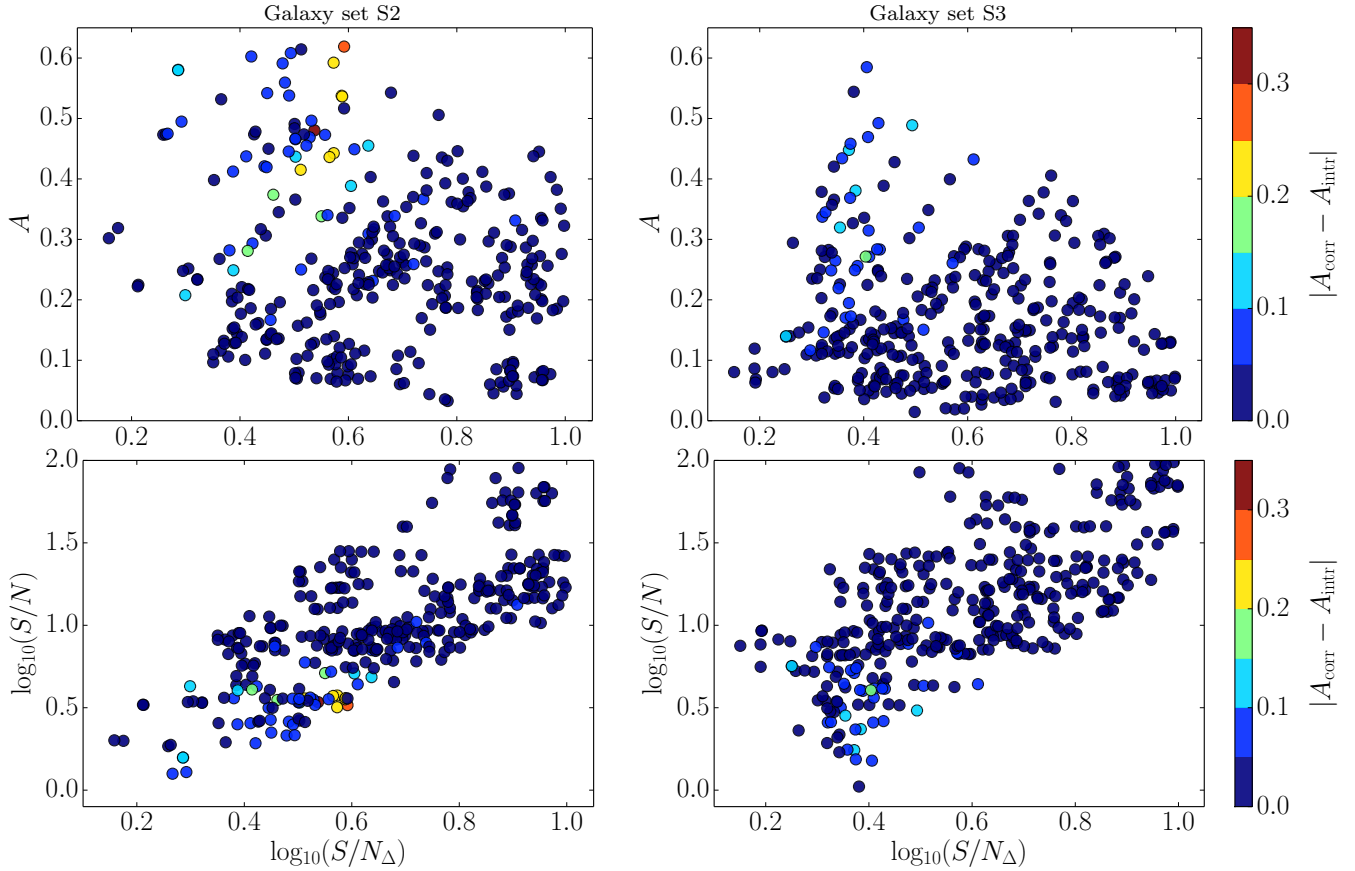


Figure 23. Quality evaluation of the A parameter correction with respect to the position in parameter space. The colours indicate to what extent the measured value can be corrected within the original value. Both test sets show sufficient correction within 0.35, the majority of objects can be corrected within 0.1 of the original value. There is no trend for the correction quality in the 3D parameter space used for correction.

WHISP or Atlas^{3D}, which contain galaxies in different environments. This will be the subject of a forthcoming paper.

Eventually, the asymmetry parameter should be extended to 3D which would allow for the use of the full information contained in a data cube. In contrast to detailed 3D modelling (e.g. with TIRIFIC), this approach would represent an alternative for the estimation of a 3D asymmetry for objects with low angular resolution.

ACKNOWLEDGMENTS

NG and JmvdH acknowledge support from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement nr. 291531.

We would like to thank the referee for his/her detailed and useful comments and suggestions.

REFERENCES

Abraham R. G., Tanvir N. R., Santiago B. X., Ellis R. S., Glazebrook K., van den Bergh S., 1996, MNRAS, 279, L47
 Abraham R. G., van den Bergh S., Nair P., 2003, ApJ, 588, 218
 Barnes D. G., et al., 2001, MNRAS, 322, 486
 Begeman, K. G. 1987, Ph.D. Thesis, University of Groningen

Bershady M. A., Jangren A., Conselice C. J., 2000, AJ, 119, 2645
 Conselice C. J., 2003, ApJS, 147, 1
 Conselice C. J., Bershady M. A., Jangren A., 2000, ApJ, 529, 886
 Di Teodoro E. M., Fraternali F., 2015, MNRAS, 451, 3021
 Duffy A. R., Meyer M. J., Staveley-Smith L., Bernyk M., Croton D. J., Koribalski B. S., Gerstmann D., Westerlund S., 2012, MNRAS, 426, 3385
 Giovanelli R., et al., 2005, AJ, 130, 2598
 Holwerda B. W., Pirzkal N., de Blok W. J. G., Bouchard A., Blyth S.-L., van der Heyden K. J., Elson E. C., 2011a, MNRAS, 416, 2401
 Holwerda B. W., Pirzkal N., de Blok W. J. G., Bouchard A., Blyth S.-L., van der Heyden K. J., Elson E. C., 2011b, MNRAS, 416, 2415
 Holwerda B. W., Pirzkal N., Heiner J. S., 2012, MNRAS, 427, 3159
 Holwerda B. W., et al., 2014, ApJ, 781, 12
 Jog C. J., Combes F., 2009, PhR, 471, 75
 Johnston S., et al., 2008, ExA, 22, 151
 Józsa G. I. G., Kenn F., Klein U., Oosterloo T. A., 2007, A&A, 468, 731
 Kerp J., Winkel B., Ben Bekhti N., Flöer L., Kalberla P. M. W., 2011, AN, 332, 637
 Lisker T., 2008, ApJS, 179, 319-325
 Lotz J. M., Primack J., Madau P., 2004, AJ, 128, 163
 Martinsson T. P. K., 2011, Ph.D. Thesis, University of Groningen
 Noordermeer E., van der Hulst J. M., Sancisi R., Swaters R. A., van Albada T. S., 2005, A&A, 442, 137

- Noordermeer E., 2006, Ph.D. Thesis, University of Groningen
- Oosterloo T., Verheijen M., van Cappellen W., "The latest on Apertif", in proceedings of "ISKAF2010 Science Meeting", PoS(ISKAF2010)043
- Rogstad D. H., Lockhart I. A., Wright M. C. H., 1974, *ApJ*, 193, 309
- Sancisi R., Fraternali F., Oosterloo T., van der Hulst T., 2008, *A&ARv*, 15, 189
- Scarlata C., et al., 2007, *ApJS*, 172, 406
- Sellwood J. A., Sánchez R. Z., 2010, *MNRAS*, 404, 1733
- Serra P., et al., 2012b, *MNRAS*, 422, 1835
- Serra, P., Jurek, R., & Flöer, L. 2012a, *PASA*, 29, 296
- Serra P., et al., 2015, *MNRAS*, 448, 1922
- Spekkens K., Sellwood J. A., 2007, *ApJ*, 664, 204
- Swaters R. A., van Albada T. S., van der Hulst J. M., Sancisi R., 2002, *A&A*, 390, 829
- van der Hulst J. M., van Albada T. S., Sancisi R., 2001, *ASPC*, 240, 451
- van der Hulst J. M., Terlouw J. P., Begeman K. G., Zwitter W., Roelfsema P. R., 1992, *ASPC*, 25, 131
- van Eymeren J., Jütte E., Jog C. J., Stein Y., Dettmar R.-J., 2011, *A&A*, 530, AA30
- Verheijen M. A. W., Oosterloo T. A., van Cappellen W. A., Bakker L., Ivashina M. V., van der Hulst J. M., 2008, *AIPC*, 1035, 265
- Walter F., Brinks E., de Blok W. J. G., Bigiel F., Kennicutt R. C., Jr., Thornley M. D., Leroy A., 2008, *AJ*, 136, 2563
- Warner, P. J., Wright, M. C. H., & Baldwin, J. E. 1973, *MNRAS*, 163, 163

APPENDIX A: APPENDIX

A1 TIRIFIC input

Fig. A1 shows some example TIRIFIC input parameters for the galaxy models displayed in Fig. 11. The model consists of 11 rings. The input parameters include the rotation curve, the inclination and position angle, the surface brightness profile, the scale height and the parameters that regulate the lopsidedness. For the plot in Fig. 11 the model was re-gridded to a resolution of 7 beams across the major axis.

This paper has been typeset from a $\text{\TeX}/\text{\LaTeX}$ file prepared by the author.

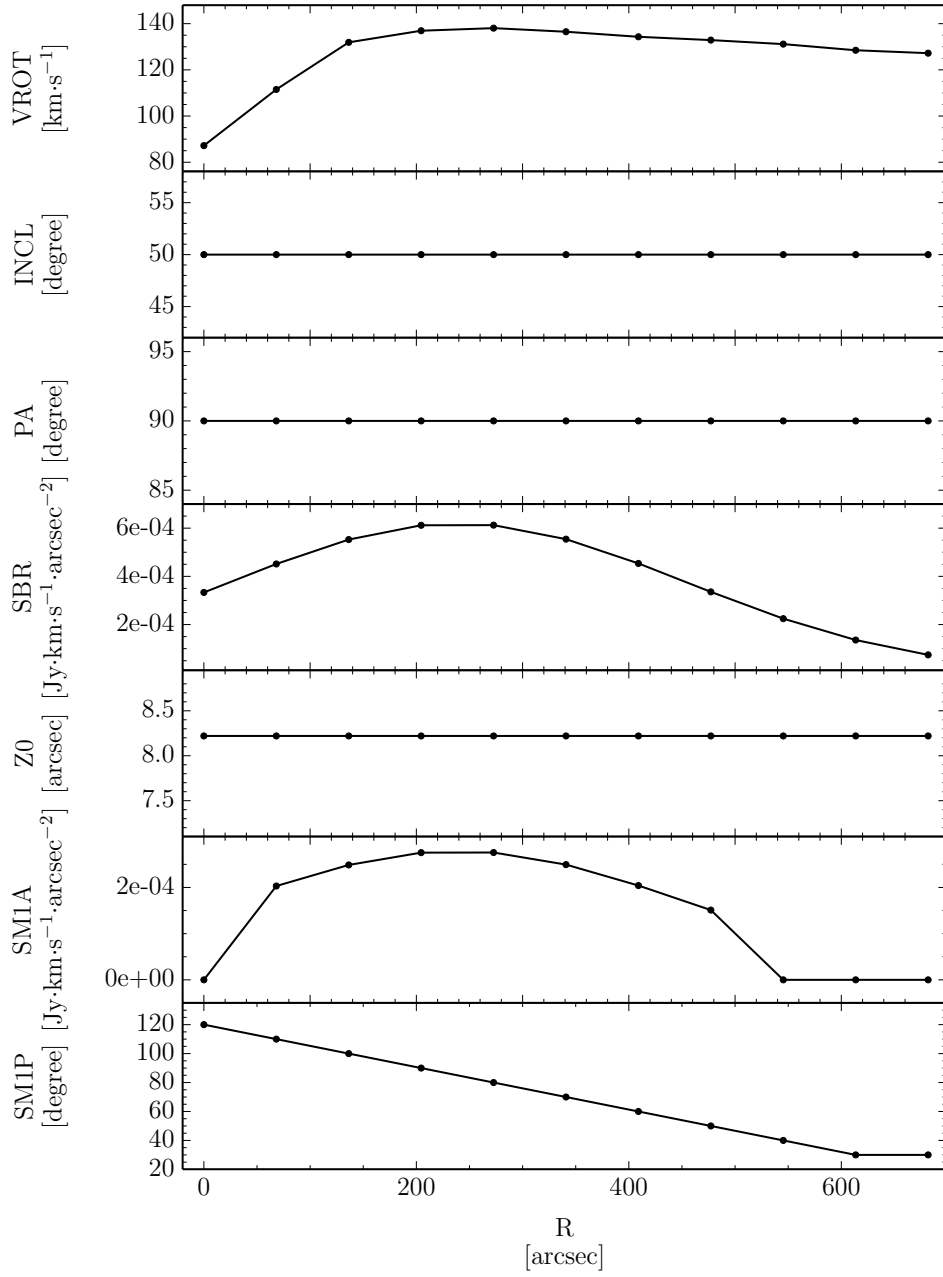


Figure A1. Example model input parameters for TiRiFiC. Models consist of 11 rings. Every ring is characterised by a rotational velocity (VROT), an inclination (INCL), a position angle (PA), a surface brightness (SBR), a scale height (Z0) and the amplitude (SM1A) and phase (SM1P) of the first order harmonic surface brightness distortion.